American Journal of Medical Genetics 92:212–219 (2000)

Major Gene Segregation of Actinic Prurigo Among North American Indians in Saskatchewan

Audrey H. Schnell,[1]* Robert C. Elston,[1] Peter R. Hull,[2] and Peter R. Lane[2]

[1]*Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHealth Campus, Case Western Reserve University, Cleveland, Ohio*

[2]*Division of Dermatology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada*

Actinic prurigo is an idiopathic, familial photodermatosis seen especially in American Indians. Segregation analysis was performed on 12 Saskatchewan pedigrees with American Indian ancestry, comprising a total of 1,148 individuals, ascertained via probands diagnosed with actinic prurigo. Although a high degree of familial aggregation has been noted in the past and dominant inheritance has been suggested, no formal segregation analysis has been attempted. Actinic prurigo has a variable age of onset and, therefore, age at the time of censoring must be taken into account in the analysis. However, as these ages of 57% of the unaffected individuals were missing, an algorithm was devised to impute the missing ages from known birth years in the family based on the age differences among relatives and spouses. Using these imputed ages, simple dominant inheritance with incomplete penetrance and a single age of onset distribution was found. The method for imputing the ages at examination was evaluated, as was the correction for ascertainment, by using alternative methods and comparing the results. Regardless of the method used, a dominant mode of inheritance without any multifactorial component remained the best hypothesis. Am. J. Med. Genet. 92:212–219, 2000.

KEY WORDS: age of onset; ascertainment; dominant inheritance; polymorphous light eruption

INTRODUCTION

Actinic prurigo (AP) belongs to the group of idiopathic photodermatoses that includes polymorphous light eruption (PLE), hydroa vacciniforme, and solar urticaria. There are many reports of AP in American Indians from Columbia [Londoño et al., 1968], Canada [Birt and Davis, 1975; Lane et al., 1992], Mexico [Hojyo-Tomoka et al., 1997], and the United States [Fusaro and Johnson, 1986]. AP also has been described in European Caucasians [Addo and Frain-Bell, 1984] and in Asiatics [Tham and Tay, 1992/1993]. Although early North American cases of AP were reported as hereditary PLE [Fusaro and Johnson, 1986], and some authors have preferred to refer to AP as a variant of PLE [Epstein, 1980], Lane et al. [1991] have indicated that AP can be distinguished from PLE both clinically and by HLA typing. Grabczynska et al. [1999] were able to demonstrate distinct demographic and clinical features distinguishing the two conditions in 90% of cases. In this study from the United Kingdom, where it is common and affects 20% of the Caucasian population, PLE appeared to coexist with AP, progress to AP, or to follow AP in 35% of patients.

Some hours after sun exposure, AP patients experience an intense pruritus and develop a dermatosis on the face and exposed areas. Skin signs include erythematous weeping areas, vesicles, papules, nodules, and plaques. Excoriations are common. The lips and conjunctivae are often affected. The skin lesions may extend onto covered areas. Clearing occurs for the majority in the winter, but for some the dermatosis may persist throughout the year. An exacerbation is experienced by all patients in the summer. Of 128 AP patients reported from the Canadian province of Manitoba (adjacent to Saskatchewan), 70% had onset prior to age of 10 [Birt and Davis, 1975]. In the Saskatchewan series of 93 AP patients [Lane et al., 1992], 45% had an age of onset younger than 9 years and 72% had on onset before the age of 20 years. Females were twice as likely to be affected males in Birt's study and three times as likely to be affected in Lane's study [Lane et al., 1992].

Londoño et al. [1968] first documented familial AP. Birt and Davis [1975] found 75% of cases had a positive

Grant sponsor: National Center for Research Resources; Grant number: 1 P41 RR03655; Grant sponsor: National Institute of General Medical Sciences; Grant number: GM 28353; Grant sponsor: University of Saskatchewan.

*Correspondence to: Audrey H. Schnell, Department of Epidemiology and Biostatistics, Case Western Reserve University, 2500 MetroHealth Drive, Cleveland, OH 44109.

Received 19 August 1999; Accepted 3 February 2000

family history and suggested that the disease was transmitted as an autosomal dominant trait with incomplete penetrance. Lane et al. [1992] had similar findings. Orr and Birt [1984] examined an Inuit population, a population with no Indian ancestry, and found that 75% of cases had a positive family history. Among Caucasians, Addo and Frain-Bell [1984] found that among half of 37 Caucasian British AP patients there was evidence of familial photosensitivity.

Although there has been no formal segregation analysis of AP, several studies have focused on HLA associations in both American Indian and Caucasian AP patients. HLA evidence includes the finding in American Indians of an increased frequency of HLA-A24 [Sheridan et al., 1990] and HLA-C4 [Bernal et al., 1990]. Menagé et al. [1996] published the association of HLA-DR4 in 90% of British Caucasian AP patients, with the rare DR4 subtype DRB1*0407 in 60%. These findings were confirmed by O'Reilly et al. [1996] and Dawe et al. [1997] in Caucasians, and by Hojyo-Tomoka et al. [1997] in American Indians.

MATERIALS AND METHODS

Subjects

Twelve pedigrees were compiled from 20 probands with AP, diagnosed by dermatologists in Saskatchewan clinics between 1975 and 1993. Ten patients from these 12 families were included among the 32 HLA patients reported by Sheridan et al. [1990]. Of the 12 families, 7 were Cree, 4 were Cree Métis (part Cree, part Caucasian), and 1 was Sioux Métis (part Sioux, part Caucasian). From our previous study we know that the clinical presentation of AP is the same in both Indians and Métis; thus when we refer to "American Indians" we include both Indians and Métis [Lane et al., 1992]. Each proband had been referred by a family practitioner to a dermatologist for diagnosis. In 5 families there was more than one proband. In 2 of these families there were 2 probands and in 3 families there were 3. Pedigree data were collected in 1993–1994. Many of these patients live in remote areas of Saskatchewan and information on family members was often difficult to obtain. Not all family members, affected or unaffected, were seen by a dermatologist. However, this population is familiar with the symptoms of AP; therefore, informant reports of whether other family members were affected or not affected were deemed to be reliable. Each subject's age, in years at the time of data collection for living subjects and at the time of death for deceased subjects, was recorded. For unaffected family members, this is the latest age at which they are known to be unaffected, and for affected family members this is the age by which they are affected. Thus, we call this age at the time of censoring the "age at examination" [Elston, 1973].

There was a total of 1,148 individuals in these 12 families in whom affection status was known. Pedigrees ranged in size from 28 to 177 people, with a mean pedigree size of 97. There were 3 to 6 generations per pedigree, with an average of 4.5. Table I gives a breakdown of pedigree members by disease status. There were 493 (43%) males, 503 (44%) females, and 152 (13%) for whom sex was unknown. Of the 1,148 individuals, 101 were affected with AP and of these, 38 (38%) were male and 63 (62%) were female. Age of onset was known in 86 (86%) cases. Age of onset ranged from 1 to 57 years, with a mean of 12.22 years (SD = 11.92 years). When age of onset information for an affected individual was unavailable, the age at examination was used in the analysis. Of the 15 affected individuals who had an unknown age of onset, 5 had a known age at examination, and for 9 other individuals it was possible to impute an age at examination as discussed later. For one affected individual, age of onset was unknown and age at examination was neither available nor could be imputed; this individual was removed from the analysis.

There were 1,047 unaffected individuals. Of these, 455 (43%) were male, 440 (42%) were female, and 152 (14.5%) had unknown sex. Age at examination was known for 453 (43%) individuals and their mean age was 18.64 years (SD = 15.38 years). Age at examination was unknown for 594 (57%) individuals due to either missing year of birth (590 cases) or missing year of death (4 cases). Age at examination could be imputed for 526 of these 594 individuals by the method described below. Missing information on year of birth occurred most often in earlier generations. The percentages of data for which year of birth was available were 11, 36, 39, 55, 76, and 75% for the 1st through 6th generations, respectively.

Methods

Segregation analysis. Segregation analysis was performed using the REGTL program that is part of the Statistical Analysis for Genetic Epidemiology (S.A.G.E.) computer package [S.A.G.E., 1998]. REGTL performs maximum likelihood segregation analysis of a dichotomous trait with a variable age of onset that follows, possibly after transformation, a logistic distribution. The models in REGTL allow for up to 3 "types" of individuals (AA, AB, and BB), where type refers to the presence of some factor (A and B) that can be transmitted from generation to generation. Type is defined in terms of transmission: 2 people are of the same type if and only if their offspring by a mate of given type have the same phenotypic distribution. The probability that this factor is transmitted from parent to offspring is a transmission probability that depends on the parent's type: the probability that a person of a given type

TABLE I. Sex and Age Information Available by Affection Status*

	Affected	Unaffected
Total number	101	1047
Male	38	455
Female	63	440
Sex unknown	0	152
Age of onset known	86	NA
Age at examination known	5[a]	453
Age at examination imputed	9[a]	526
Age unknown and not imputed	1[a]	68

*NA, not applicable.
[a]with unknown age of onset.

transmits factor A to offspring. Mendelian inheritance, if it occurs, is assumed to be through a single autosomal locus with 2 alleles (A and B), where A is the allele associated with the disease. The word *type* is used generally, whatever the mode of transmission, Mendelian inheritance being a specific mode of transmission in which the types are genotypes and the transmission probabilities for the three genotypes (AA, AB, and BB) are 1, .5, and 0, respectively. Assume that there is a proportion of people in the population who are susceptible to the disease and a proportion who are not susceptible, for whom age of onset is therefore irrelevant. Let susceptibility, γ, be the probability that a person is susceptible. The likelihood for an individual who is affected at age a is taken to be γ f(a), that for an individual affected by age at examination a′ (if age of onset is unknown) to be γ F (a′), and for an individual who is unaffected by age a′ to be 1- γ F(a′), where f is the logistic density and F is the corresponding cumulative distribution. The density function is thus of the form

$$f(a) = \frac{\alpha \exp\{\beta + \alpha a + \text{familial effects}]}{1 + \exp[\beta + \alpha a + \text{familial effects}]^2} \qquad (1)$$

and the cumulative distribution is

$$F(a) = \frac{1}{1 + \exp\{ -[\beta + \alpha a + \text{familial effects}]\}} \qquad (2)$$

where β is a baseline parameter, α is the age coefficient, the mean of the distribution conditional on familial effects is $-(\beta$ + familial effects)/α, and the variance of the conditional distribution is $\pi^2/3\alpha^2$. The familial effects include spouse, parental, and sibling effects, quantified by regressive coefficients in order to model multifactorial transmission [Bonney, 1986]. Major effect transmission is modeled by the 3 transmission probabilities τ_u (u = AA, AB, or BB), the probability that a person of type u transmits A to an offspring. If Hardy-Weinberg equilibrium is assumed, then the relative frequencies of the 3 types of individuals in the population are q^2, 2q(1-q), and $(1\text{-}q)^2$ for AA, AB, and BB, respectively, so that q is the frequency of allele A when there in Mendelian transmission.

Two different methods for modeling how types determine disease are available in REGTL. In the first method (Model 1), types (or genotypes in the case of Mendelian inheritance) influence the age of onset distribution. Here all 3 types of individuals have the same susceptibility to disease γ, and there can be a separate baseline parameter β associated with each type (β_{AA}, β_{AB}, β_{BB}), which can also be made to depend on sex. There may be a common age coefficient α, possibly sex dependent, or α can also be made type (and sex) dependent. Alternately, in the second method (Model 2), type influences susceptibility. Therefore, there are up to 3 different susceptibilities (γ_{AA}, γ_{AB}, γ_{BB}), but a single age of onset distribution for all individuals so that the parameters β and α are common to all types, though possibly sex dependent. When there is only one type, Model 1 and Model 2 are necessarily equivalent. Again, these types correspond to genotypes in the case of Mendelian inheritance.

Hypotheses without transmission from generation to generation, but allowing for 1, 2, or 3 age-of-onset distributions dependent on type (Model 1), were first compared using commingling analysis. In REGTL this is accomplished by restricting the 3 transmission probabilities to equal q. This is the same as assuming that the factor responsible for disease is not transmitted from generation to generation, but rather occurs at random with the same frequency in all generations. Even though there is no transmission and type is actually defined in terms of transmission, it is still possible (under Model 1) to have 1, 2, or 3 different age-of-onset distributions—due, for example, to some major random environmental factor ("random environmental" hypothesis). When there is only 1 distribution, the hypothesis corresponds to there being no major effect ("NME" hypothesis).

Under Model 1, assuming a common age coefficient, a dominant hypothesis corresponds to $\beta_{AA} = \beta_{AB}$, a recessive hypothesis to $\beta_{AB} = \beta_{BB}$, and for codominance β_{AA}, β_{AB}, and β_{BB} are estimated separately. For Model 2, a dominant hypothesis corresponds to $\gamma_{AA} = \gamma_{AB}$, a recessive hypothesis to $\gamma_{AB} = \gamma_{BB}$, and for codominance γ_{AA}, γ_{AB}, and γ_{BB} are estimated separately. It should be noted that for Model 2 the random environmental and NME hypotheses are mathematically identical (even though the parameter estimates that are found to maximize the likelihood may be different, because not all the parameters are estimable when the 3 transmission probabilities are equal to q). Under Model 2, which fitted the data better than Model 1, more complex hypotheses were also tested by allowing for multifactorial effects of mothers, fathers, and spouses. Hypotheses were fitted with and without simultaneously estimating a transformation parameter [Box and Cox, 1964] for the age-of-onset distribution. Several sets of initial estimates were used for each hypothesis fitted in order to find the global maximum, rather than a local maximum, of the likelihood. A general model in which all parameters were freely estimated, but with the restriction that the phenotypic distribution is homogeneous across generations, was fitted to the data and provided the baseline to which the hypotheses were compared. All hypotheses correspond to restrictions applied to the general model, and were tested on the basis of twice the difference of their $\log_e$ likelihoods from that of the general model. Additionally, 2 hypotheses were compared with each other when one could be considered as a special case of the other. Under certain conditions, the difference in $\log_e$ likelihoods is asymptotically distributed as a chi-square statistic (χ^2_{df}) when the hypothesis holds, with degrees of freedom (df) equal to the difference in the number of independent parameters being estimated between the hypothesis and the model or between the 2 hypotheses. For testing mixtures of distributions, the regularity conditions necessary for this test to be valid do not hold; nevertheless χ^2_{df} appears (from simulation studies) to be a reasonable approximation in the upper tail of the distribution.

Imputing age at examination. Because this disease has a variable age of onset, age at examination for unaffected individuals, that is, the censoring age, must

TABLE II. Age of Mother and Father at Birth of Offspring*

Offspring		1	2	3	4	5	6	7	8	9	10	11
Number	M	90	64	50	41	29	20	12	8	6	5	3
	F	56	45	32	25	21	15	9	6	5	3	2
Mean	M	19.27	21.27	23.38	25.98	27.41	30.80	32.08	32.37	33.0	36.60	36.37
	F	77	24.22	26.53	28.24	30.95	34.80	37.33	38.50	41.80	44.00	48.00
Median	M	19	20.5	23	25	27	30.5	31	31	33	37	35
	F	22	24	26	29	32	35	38	37.5	42	42	48
SD	M	3.74	4.22	3.97	4.22	4.76	4.87	5.09	4.14	3.69	5.32	7.10
	F	5.77	4.41	4.97	5.04	4.57	4.41	4.0	4.23	3.70	6.25	7.07
Range	M	10–29	12–33	13–32	19–36	16–38	22–42	25–43	26–39	27–38	29–42	30–44
	F	16–51	17–36	19–37	20–39	23–40	28–44	33–46	35–46	37–47	39–51	43–53

*M, mother; F, father.

be taken into account in the segregation analysis. However, this information was missing for many members in each family largely due to the fact that it was difficult to interview many of the family members who lived in remote areas and often did not have access to a telephone. Therefore, we devised and evaluated an algorithm for imputing this missing information. From those pedigree members for whom ages at last birthday were known, the differences in age between spouses, between each parent and each offspring (1st offspring, 2nd offspring, etc.) and between consecutive siblings were calculated. Tables II and III contain the mean, median, and standard deviations for the calculated age differences for mothers, fathers, spouses, and siblings. The median age difference between a mother and her first offspring was estimated to be 19 years and that between a father and his first offspring 22 years. The median difference between the ages of spouses was estimated to be 3 years. For consecutive siblings, when there were less than 8 siblings a median age difference of 2 years was used in the calculation of missing age at examination, whereas if there were 8 or more siblings, 1.5 years was used because it was noted that this was the median difference in age between all consecutive siblings in the few families with 8 or more siblings (If there was an even number of observations, the median was taken as an average of the two middle observations.) Median differences were then used to impute missing ages, whenever possible from known ages of first-degree relatives, but occasionally from first-degree relatives whose ages were themselves imputed. All equally close relatives' information was used to impute the unknown ages (i.e., siblings, parents, and offspring were all used to impute an unknown age when available) and averages of these imputed ages used. For mothers, her offspring, her siblings, and her mother were used. For children, only the mother's age and those of siblings (when known) were used, and not the father's, as a mother's age is more restricted biologically. For fathers, all information, including the spouse's age, was used. If the algorithm produced an impossible result (which happened rarely), the most reasonable age was assumed. For example, if an offspring's birth year was estimated to be after 1994 (the last year data were collected) then 1994 was used. As an illustration, consider a family where the birth years of 4 offspring and mother were used to impute the birth year for the father. This was done by subtracting the median differences in age between a father and each of the first 4 offspring (22, 24, 26, and 29, Table II) from the birth year of each corresponding offspring, the median difference in spouses' age (3, Table III) from the wife's birth year, and then averaging the resulting 5 differences to obtain the father's birth year. If a mother's or father's age at death was unknown, then an age at examination was set equal to age at the birth of the last offspring as estimated from the data. This approach was used to take advantage of all possible information, as parents obviously had to be at least of an age to give birth to the number of offspring they had before their death. For example, in this data set the median mother's age at the birth of her 3rd child was estimated to be 23 years (Table II). So, if a woman was deceased and her year of death was unknown but she had 3 children, her age at examination was set to 23, the median age in this data set for a mother at the birth of her 3rd child. Therefore, age at examination in these cases was most likely set to a younger age than the actual age at death. For 12 cases (11 unaffected and 1 affected individual) no age at examination could be imputed because the individual was deceased, there was no date of death, and the individual had no offspring.

For 56 individuals (all unaffected) in 2 pedigrees, age at examination could not be imputed because in some

TABLE III. Age Differences Between Spouses and Between Consecutive Siblings

		Siblings									
	Spouses	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10	10–11
Number	22	107	72	52	37	28	15	10	8	6	4
Mean	5.18	2.66	2.75	2.40	2.41	2.43	1.93	1.90	2.38	3.50	1.50
Median	3	2	2	2	2	2	2	2	1.5	2	1.5
SD	7.06	1.92	1.69	2.41	1.28	1.50	.88	.88	2.72	2.81	.58
Range	0–33	1–14	0–7	0–11	1–6	1–8	1–3	1–3	1–9	2–9	1–2

of the generations the birth order indicated in the data became suspect during the process of imputing birth years. For individuals in the generations where birth order was suspect, an age at examination could only be determined for individuals who had offspring. This was done by using the offspring birth year(s), if known, to obtain the parent birth years; or, when the offspring birth years(s) were unknown, a minimal age was determined from the number of offspring using the same method as used for deceased individuals who had offspring. These 56 individuals had no offspring, and therefore even a minimal age at examination could not be determined for them. Thus, there were a total of 68 individuals for whom age at examination was unknown and no information could be imputed; these individuals were considered missing in all analyses.

The algorithm used to impute the missing ages at examination was first evaluated in the following way. The ages that were imputed were taken as known, the known ages were set to missing, and then the latter were imputed from the now "known" ages using the same algorithm. This resulted in a new data set, and the parameter estimates and the pattern of differences among the likelihoods were compared between the 2 data sets.

Two additional ways of evaluating the effect of imputing the missing ages were used. First, a data set was constructed in which deceased individuals who had been given ages at examination that were probably too young, based on the number of offspring they had, were now given much older ages at examination, equal to their ages as if they were still alive in 1994 but not greater than 90 years of age. This was done only for deceased individuals whose actual year of death was unknown and whose birth year was either known or had been imputed (N = 35). It was not done for the individuals in the 2 families for whom birth years could not be imputed but who had nevertheless been given an age at examination based on the number of offspring born, or for the deceased individuals who had not been given an age at examination in the original data set because they had neither a known date of death or offspring. By imputing a younger-than-probable age at examination in 1 data set and the oldest possible age at examination in another, it was possible to evaluate the effects of the algorithm using two extremes of age at examination. Second, only individuals with known ages were used to estimate parameters and obtain the likelihood, specifically for the general model and 2 of the hypotheses (the most parsimonious and that of no major effect).

Ascertainment. Because families were ascertained via proband(s), a correction for ascertainment was included. Single ascertainment can be corrected for by conditioning the likelihood on the probands being affected by their ages at examination or, if age at examination is unavailable, their ages of onset. Alternatively, it is possible to correct for ascertainment by conditioning the likelihood on the proband's actual age of onset or, if this is unavailable, the age at examination. However, although it is always valid to condition on the actual ages of onset, it is advantageous to condition on the probands being affected by their ages at examination because this retains more information. Both conditioning schemes were employed in order to check the consistency of the results.

In 5 families there was more than one proband per family. This is a concern because only in the case of single ascertainment (i.e., one proband per family) is a correction possible that yields consistent parameter estimates. For this reason, in families where there was more than one proband, initially the first (the index cases) were used as probands in the ascertainment correction, or the first 2 if seen on the same day, resulting in 16 of the 20 probands being used. This number is halfway between the entire set of probands (20) and the number of probands there would have been if there had been only one proband per family (12). To evaluate the effect of not having single ascertainment in this study, parameter estimates and likelihoods were also obtained where conditioning was performed on all 20 probands' ages at examination, and again using one randomly chosen proband per family for the multiply ascertained families.

RESULTS

Table IV shows the mean, median, and standard deviation for known ages at examination, for ages at examination in the data set where age was either known or imputed from known ages, and for the data set where all ages were imputed. The median age at examination (24 years) was the same in the 2 data sets where some or all of the ages at examination were imputed, and this was older than the median of 15 years for the known ages at examination.

Conditioning on age of onset or age at examination yielded virtually the same results in terms of parameter estimates and tests of hypotheses. Therefore, only the results from conditioning on age at examination are reported. When conditioning on age at examination, the standard deviations of the parameter estimates were slightly smaller, as would be expected because more information is retained.

For nongenetic hypotheses where age of onset is dependent on type (Model 1), 2 distributions fitted the data significantly better than 1 ($\chi^2_2 = 34.37, P < .005$), but for 3 distributions the likelihood remained virtually unchanged. Furthermore, there was an indication that the model was overparameterized: The matrix of second partial derivatives of the $\log_e$ likelihood could not be inverted because it was close to singular, indicating that too many parameters were being estimated.

TABLE IV. Comparison of the Distributions for Known, Known and Imputed, or All Imputed Ages at Examination for Unaffected Individuals

	Age known	Age known and imputed	All ages imputed
Number of subjects	453	979	974
Mean age	18.64	24.61	24.87
Median age	15	24	24
SD	15.38	16.79	16.75
Range of ages	1–86	1–98	1–98

TABLE V. Parameter Estimates and $\log_e$ Likelihoods for Model 1 Hypotheses*

Parameter	General model	Dominant	Recessive	Codominant	Random environmental	NME
q	.046	.031	.320	.033	.358	NA
β_{AA}	−.682	−3.622	−3.532	−.887	−1.39	−3.610
β_{AB}	−4.196	−3.622	−72.648	−3.941	−8.73	$=\beta_{AA}$
β_{BB}	−50.052	−50.052	−72.648	−50.052	−13.85	$=\beta_{AA}$
α	1.256	1.041	1.044	1.163	3.271	1.031
γ	.619	.732	.828	.676	.174	.207
Log_e likelihood	−528.05[a]	−530.49	−536.30	−529.54	−546.29	−558.30
χ^2	–	4.88	16.50	2.98	36.48	60.50
df	–	3	3	2	2	5
P	–	NS	<.01	NS	<.001	<.001

*$\log_e$ age of onset.
[a]τ_{AA} and τ_{BB} go to a lower bound of 0. See text for definition of all symbols.

The χ^2_1 for 1, 2, and 3 distributions, comparing the likelihoods with and without a transformation of age, were 67.57, 46.47, and 55.60, respectively. Thus, simultaneously estimating a transformation for age significantly improved the fit. In all cases, the power parameter (λ_1) was restricted to being positive and was estimated to be 0, indicating that a log transformation of age was appropriate. In view of these results, age was log transformed for all further analysis. With a transformation, 2 distributions again fitted significantly better than 1 ($\chi^2_2 = 14.20$, $P < .005$) and 3 distributions fitted significantly better than 2 ($\chi^2_1 = 9.87$, $P < .01$). When compared to the general model (Table V), the no major effect hypothesis (NME) could be rejected ($\chi^2_5 = 60.50$, $P < .001$), as could the random environmental hypothesis ($\chi^2_2 = 36.48$, $P < .001$) and a recessive hypothesis ($\chi^2_3 = 16.50$, $P < .001$). A dominant hypothesis, on the other hand, fitted the data well ($\chi^2_3 = 4.88$) and codominance did not fit the data significantly better ($\chi^2_1 = 1.90$). Assuming dominance, the estimated gene frequency was .031 and the estimated susceptibility was .732, indicating that the disease is not fully penetrant. In Table V and subsequent tables, the estimates of the parameter α are appropriate for $\log_e$ age whereas the coefficients reported by REGTL are standardized by their geometric mean.

However, too many parameters were being estimated in Model 1, as the matrix of second partial derivatives could not be inverted. For this reason the second method of modeling the disease, where there is a single age-of-onset distribution and susceptibility depends on type (Model 2), is preferable, because there were fewer parameters that needed to be estimated. The $\log_e$ likelihoods for the 2 general models differed by less than 2.0. The same hypotheses that were tested under Model 1 were again tested by comparison to the general model with arbitrary transmission probabilities and homogeneity across generations (Table VI). Again, as under Model 1, recessive ($\chi^2_3 = 12.60$, $P < .01$) and NME ($\chi^2_5 = 56.61$, $P < .001$) hypotheses could be rejected. (Under Model 2 the random environmental hypothesis and the NME hypothesis are identical.) The dominant hypothesis could not be rejected ($\chi^2_3 = 1.00$) and codominance did not fit the data significantly better ($\chi^2_1 = .14$). Assuming dominance, the gene frequency was estimated to be the same under both models; the susceptibilities γ_{AA} and γ_{AB} under Model 2 were the same as the susceptibility under Model 1, whereas the mean age of onset under Model 2 was the same as the mean age of onsets for genotypes AA and AB under Model 1. The mean age of onset was estimated to be 32.48 with a standard error (estimated by numerical double differentiation of the likelihood) of 4.58.

The addition of a multifactorial effect of having an affected spouse and/or an affected parent did not significantly improve the fit of Model 2 after allowing for a major gene ($\chi^2_3 = 6.37$), whereas multifactorial inheritance alone, without allowing for a major gene, led to a highly significant worse fit ($\chi^2_1 = 46.84$). A data

TABLE VI. Parameter Estimates and $\log_e$ Likelihoods for Models 2 Hypotheses*

Parameter	General model	Dominant[a]	Recessive	Codominant	NME
q	.032	.031 (.014)	.320	.031	NA
β	−3.612	−3.622 (.347)	−3.53	−3.62	−3.611
α	1.042	1.040 (.186)	1.044	1.044	1.030
γ_{AA}	.705	.732 (.224)	.828	1.00	.207
γ_{AB}	.724	.732 (.224)	0	.7126	$=\gamma_{AA}$
γ_{BB}	0	0	0	0	$=\gamma_{AA}$
Log_e likelihood	−529.99[b]	−530.49	−536.29	−530.42	−558.30
χ^2	–	1.00	12.60	.86	56.61
df	–	3	3	2	5
P	–	NS	*P* < .01	NS	<.001

*$\log_e$ age of onset.
[a]Numbers in parentheses for the dominant hypothesis are standard errors.
[b]τ_{AA} and τ_{BB} go to a lower bound of 0. See text for definition of all symbols.

TABLE VII. Parameter Estimates and $\log_e$ Likelihoods for Models and Hypotheses Where (1) All Ages of Examination Were Imputed With a Downward Bias, and (2) the Imputed Ages Had an Upward Bias*

	(1) All ages imputed			(2) Older ages for decreased individuals		
Parameters	General	Dominant	NME	General	Dominant	NME
q	.031	.031	NA	.031	.030	NA
β	−3.587	−3.595	−3.595	−3.608	−3.610	−3.624
α	1.039	1.038	1.025	1.043	1.044	1.023
γ_{AA}	.774	.724	.206	.704	.703	.206
γ_{AB}	.714	.724	$=\gamma_{AA}$	.698	.703	$=\gamma_{AA}$
γ_{BB}	0	0	$=\gamma_{AA}$	0	0	$=\gamma_{AA}$
Log_e likelihood	−530.59[a]	−531.10	−558.93	−529.36[a]	−529.89	−557.53
χ^2	–	1.03	56.68	–	1.06	56.34
df	–	3	5	–	3	5
P	–	NS	<.001	–	NS	<.001

*$\log_e$ age of onset.
[a]τ_{AA} and τ_{BB} go to a lower bound of 0. See text for definition of all symbols.

set with only individuals whose sex was known was used to estimate sex-dependent baseline parameters and sex-dependent susceptibilities, but neither significantly improved the fit (data not shown).

The results presented in Table VI are for a data set in which the ages at examination for deceased individuals were probably underestimated. The left half of Table VII gives results for the general model, as well as for the best fitting dominant and no major effect hypotheses, applied to the data set where all ages were imputed (which would similarly have age at examination somewhat underestimated) whereas the right half of Table VII gives the corresponding results for the data set where deceased individuals were given an older age at examination. The results in each case were virtually the same as those in Table VI: The dominant hypothesis was not rejected and the parameter estimates were almost identical. When only known ages at examination were used, the dominant hypothesis yielded a slightly larger $\log_e$ likelihood than either the no major effect or recessive hypotheses but discrimination between these hypotheses was less clear (data not shown).

Finally, the results obtained when the likelihood was conditioned on all probands, one proband per family or the subset of 16 probands were similar (Table VIII), with the largest effect being on the gene frequency and susceptibility parameters, in accordance with expectation for large pedigrees [Elston, 1979]. Under all 3 different corrections for ascertainment, the no major effect hypothesis, but not the dominant hypothesis, was rejected when compared to a general model.

DISCUSSION

Prior to this analysis, no formal segregation analysis of AP had been performed. The results found here indicate that the data are consistent with a dominant mode of inheritance with incomplete penetrance, but with no multifactorial component. A recessive mode of inheritance as well as only a multifactorial effect or a random environmental effect were all rejected. This is consistent with the literature where a dominant mode of inheritance has been proposed. In all of the general models τ_{AA} and τ_{BB} became fixed to 0 during the process of maximization. In a sample of this size, however, even using the largest estimated gene frequency, there were at most 3 people estimated to have genotype AA and the estimate of τ_{AA} must therefore be poor; τ_{AB} was estimated to be about .52, close to its Mendelian value, in each of the general models. An increase in the frequency of AP in females was reported in the literature [Lane, 1995], and the proportion of females with the disease was greater than that of males in this study. Although the estimate of the female susceptibility of the dominant genotype was found to be larger than the corresponding male susceptibility (.784 vs. .642), this difference was not significant.

Because the censoring age was missing for approximately half of the individuals, we devised an algorithm to impute it based on the known ages of close relatives. Initially, it was suspected that there might be a cohort effect for mothers such that mothers born in earlier decades would give birth to their first offspring at an earlier age than mothers born in later decades. This possibility was examined by plotting the mother's age at the birth of her first offspring against her year of birth. No such trend was observed, thus the information from all mothers was grouped together.

The mean of the known ages at examination was smaller than the mean age of examination in either data set where some or all the ages were imputed (Table V). This is not surprising, as it was in the older generations that ages were most often missing. There was no way to know with complete certainty the accu-

TABLE VIII. Parameter Estimates for Dominant Inheritance: 3 Different Corrections for Ascertainment, Model 2*

	Conditioning the likelihood on		
	All probands	16 probands	1 proband/ family
q	.025	.031	.032
β	−3.568	−3.622	−3.593
α	1.050	1.040	1.042
γ_{AA}	.679	.732	.750
γ_{AB}	.679	.732	.750
γ_{BB}	0	0	0
Log_e likelihood	−522.40	−530.49	−539.58
χ^2	.03	.99	.88
df	3	3	3
P	NS	NS	NS

*See text for definition of all symbols; $\log_e$ age of onset.

racy of the method used to impute missing birth years, but the standard deviations and ranges shown in Tables II and III give an indication. Furthermore, three different strategies were employed to discover the effect of imputing the ages. Imputing either younger or older ages for deceased individuals had no effect on the choice of the best fitting hypothesis and little effect on the parameter estimates, confirming the robustness of the tests and estimates. Although it is possible to use only those individuals with known ages in the analysis, there was then much less information to discriminate among the different hypotheses.

A further difficulty in this study was that there was more than one proband per family in 5 of the families. The effect of variously allowing for this was largest on the estimate of gene frequency and susceptibility, but did not effect the choice of the best fitting hypothesis. The gene frequency was estimated to be smaller when conditioning the likelihood on all probands and larger when conditioning on only 12 probands, as expected. The same pattern was seen for the susceptibilities: γ_{AA} and γ_{AB} were larger when conditioning on one proband per family and smaller when all the probands were used. Nevertheless, there was no effect on γ_{BB}, which was estimated to be 0 under all 3 corrections for ascertainment. All 3 conditioning strategies led to the same conclusion that dominant inheritance best fits the data. Even though the ages at examination had to be imputed for over half the sample and there was more than one proband per family, the consistency of the results increases confidence in the findings and motivates a linkage study to find the gene responsible for this disease.

ACKNOWLEDGMENTS

The results of this paper were obtained by using the program package S.A.G.E., which is supported by a U.S. Public Health Service Resource Grant (1 P41 RR03655) from the National Center for Research Resources. The study was supported in part by U.S. Public Health Service research grant GM 28353 from the National Institute of General Medical Sciences and from grants from the Clinical Research and Teaching Fund of the University of Saskatchewan. We would also like to thank the following University of Saskatchewan students, Katherine Lockhart, Zepnep Akin, and Coby Demkew-Bartell, for their assistance in the survey.

REFERENCES

Addo HA, Frain-Bell W. 1984. Actinic prurigo—a specific photodermatosis? Photo-Dermatol 1:119–128.

Bernal JE, Duran de Rueda MM, Ordonez CP, Duran C, de Brigard D. 1990. Actinic prurigo among the Chimila Indians in Columbia: HLA studies. J Am Acad Dermatol 22:1049–1051.

Birt AR, Davis RA. 1975. Hereditary polymorphic light eruption of American Indians. Int J Dermatol 14:105–111.

Bonney GE. 1986. Regressive logistic models for familial disease and other binary traits. Biometrics 42:611–625.

Box GEP, Cox DR. 1964. An analysis of transformations (with discussion). J R Statist Soc B 26:211–252.

Dawe RS, Collins P, Ferguson J, O'Sullivan A. 1997. Actinic prurigo and HLA DR4. J Invest Dermatol 108:233–234.

Elston RC. 1973. Ascertainment and age of onset in pedigree analysis. Hum Hered 23:105–112.

Elston RC. 1979. Likelihood models in human quantitative genetics. In: Sing CF, Skolnick M, editors. The genetic analysis of common diseases: applications to predictive factors in coronary heart disease. New York: Alan R. Liss. p 391–405.

Epstein JH. 1980. Polymorphous light eruption. J Am Acad Dermatol 3: 329–343.

Fusaro RM, Johnson JA. 1986. Hereditary polymorphic light eruption of the American Indian. Dermatol Clin 4:1–5.

Grabczynska SA, McGregor JM, Kondeatis E, Vaughan RW, Hawk JL. 1999. Actinic prurigo and polymorphic light eruption: common pathogenesis and the importance of HLA-DR4/DRB1 0407. Br J Dermatol 140:232–236.

Hojyo-Tomoka T, Granados J, Vargas-Alarcón G, Yamamoto-Furusho JK, Vega-Memije E, Cortés-Franco R, Flores O, Teixeira F, Domínugez-Soto L. 1997. Further evidence of the role of HLA-DR4 in the genetic susceptibility to actinic prurigo. J Am Acad Dermatol 36:935–937.

Lane PR. 1995. Actinic prurigo. In: Demis DJ, editor. Clinical dermatology. Philadelphia: Lippincott-Raven. Vol 4. Unit 19–3B. p 1–9.

Lane PR, Hogan DJ, Martel MJ, Rheeder B, Irvine J. 1992. Actinic prurigo: clinical features and prognosis. J Am Acad Dermatol 26:683–692.

Lane PR, Sheridan DP, Hogan DJ, Moreland A. 1991. HLA typing in polymorphous light eruption. J Am Acad Dermatol 24:570–573.

Londoño F, Muvdi F, Giraldo F, Rueda LA, Caputo A. 1968. Familial actinic prurigo. Dermatol Ibero Lat Am 111:61–71.

Menagé H du P, Vaughan RW, Baker CS, Page G, Proby CM, Breathnach SM, Hawk JLM. 1996. HLA-DR4 may determine expression of actinic prurigo in British patients. J Invest Dermatol 106:362–364.

O'Reilly FM, Spencer S, Darke C, Murphy GM. 1996. HLA-DR 4B1, 0407—strong association with actinic prurigo in Ireland. Br J Dermatol 135(Suppl 47):65.

Orr PH, Birt AR. 1984. Hereditary polymorphic light eruption in Canadian Inuit. Int J Dermatol 23:472–475.

S.A.G.E. 1998. Statistical analysis for genetic epidemiology, release 3.1. Computer program package available from the Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHealth Campus, Case Western Reserve University, Cleveland.

Sheridan DP, Lane PR, Irvine J, Martel MJ, Hogan DJ. 1990. HLA typing in actinic prurigo. J Am Acad Dermatol 22:1019–1023.

Tham SN, Tay YK. 1992/1993. Actinic prurigo, tropical (South-East Asian) variant. Photodermatol Photoimmunol Photomed 9:225–228.